RSAC 2025 Conference

Many Voices.
One Community.

SESSION ID: LAB3-T09

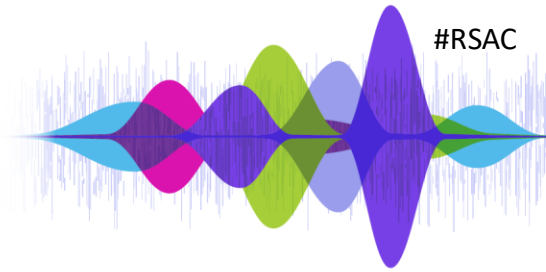
Ghost in the Machine: Deepfake Creation and Countermeasures

Bri Frost

Director of Curriculum, Security and IT Ops
Pluralsight
www.linkedin.com/in/bri-frost/

Aaron Rosenmund

Senior Director of Tradecraft and Programs
OnDefend
www.linkedin.com/in/aaronrosenmund/

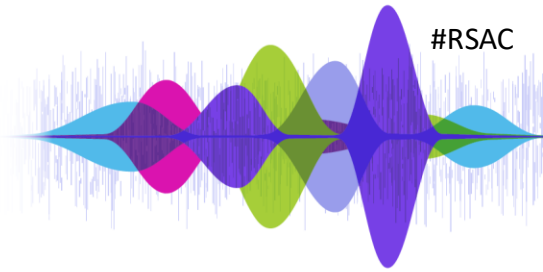


Disclaimer

Presentations are intended for educational purposes only and do not replace independent professional judgment. Statements of fact and opinions expressed are those of the presenters individually and, unless expressly stated to the contrary, are not the opinion or position of RSA Conference LLC or any other co-sponsors. RSA Conference LLC does not endorse or approve, and assumes no responsibility for, the content, accuracy or completeness of the information presented.

Attendees should note that sessions may be audio- or video-recorded and may be published in various media, including print, audio and video formats without further notice. The presentation template and any media capture are subject to copyright protection.

© 2025 RSA Conference LLC or its affiliates. The RSAC and RSAC CONFERENCE logos and other trademarks are proprietary. All rights reserved.



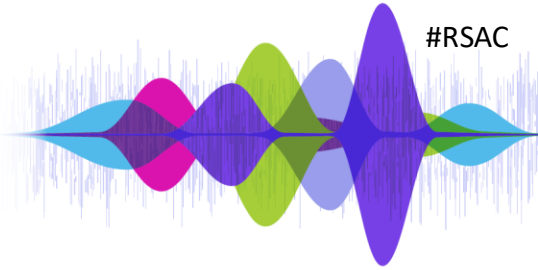
Workshop Outline

- Introduction – Deepfakes, explained
  - Can you spot the deepfake?
  - Demo: Deepfake creation live
- Lab Part One: Creating Your Own Deepfake
- Lab Part Two: Compare Your Deepfake to Authentic Video
- Lab Part Three: Can an AI-Detection Tool catch your deepfake?
- Conclusion – Mitigating Deepfake Threats in the future

Is AI a Threat?

Does AI Create Cyber Attacks?

What AI does NOT do:



- Create new attacks
- Write Zero Days or Exploits
- Gain access to a network
- Create a vulnerable code in an application

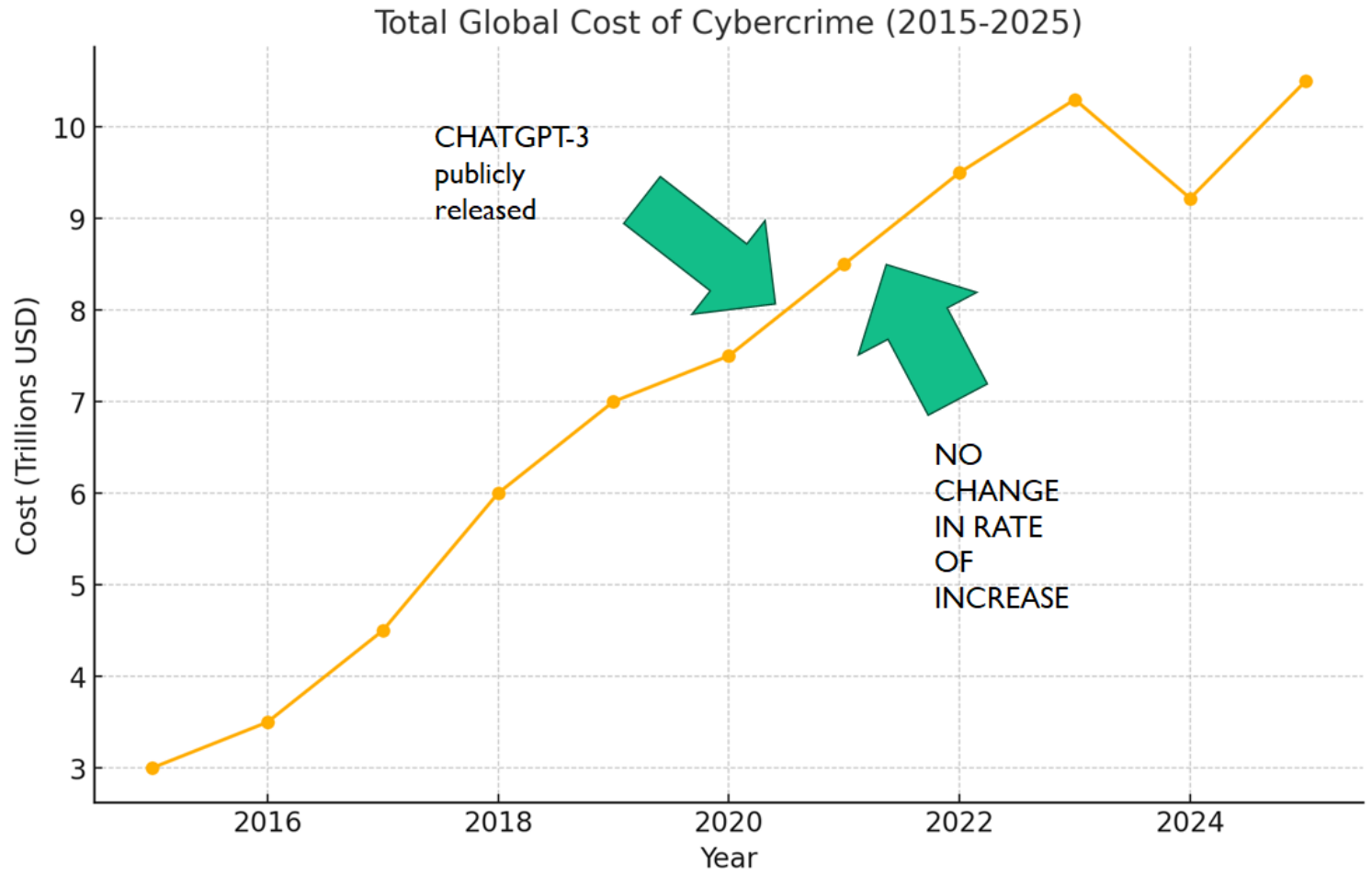
Humans do these things – not AI.

RSAC 2025 Conference

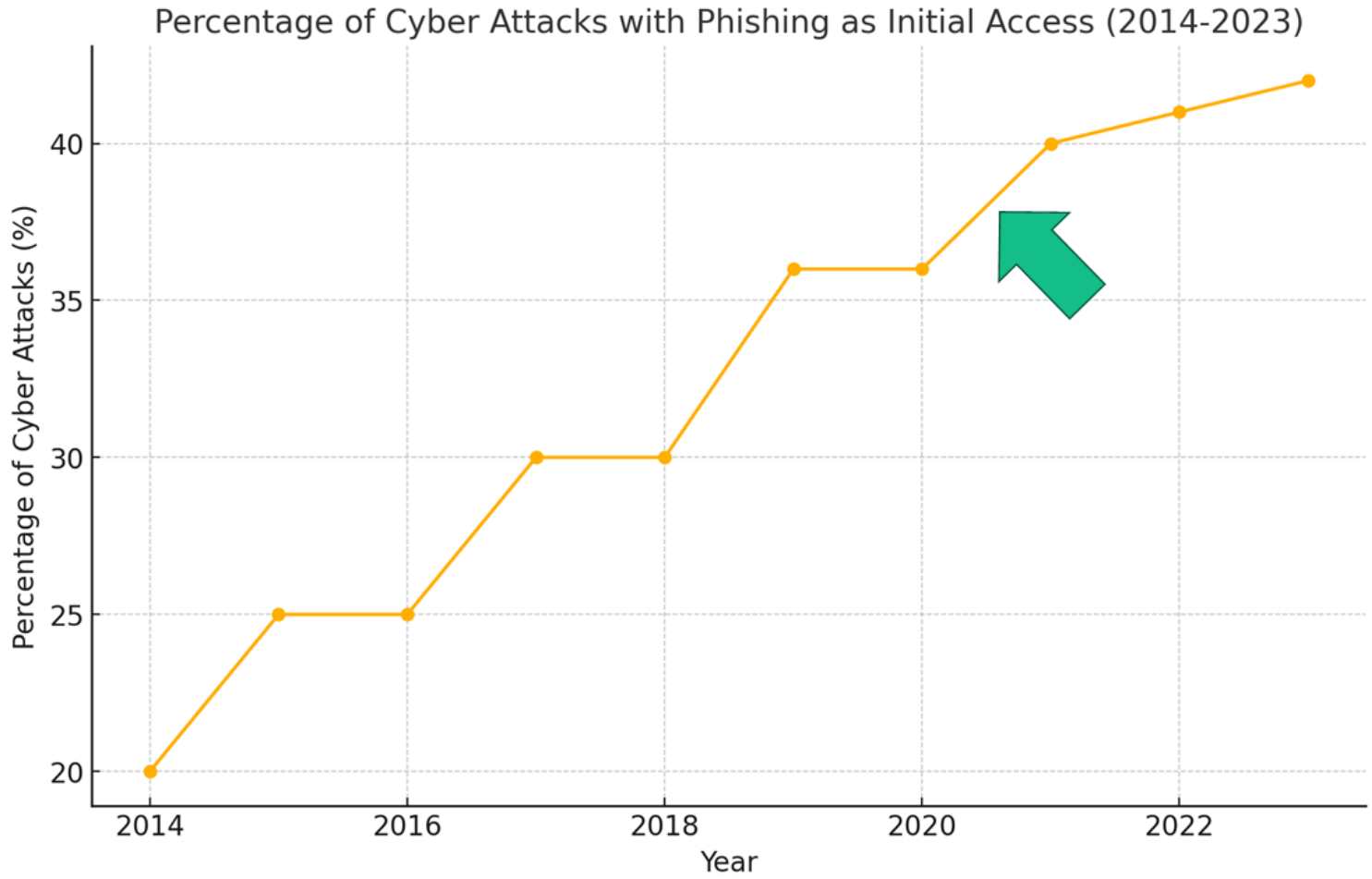
Many Voices.
One Community.

Let us prove it

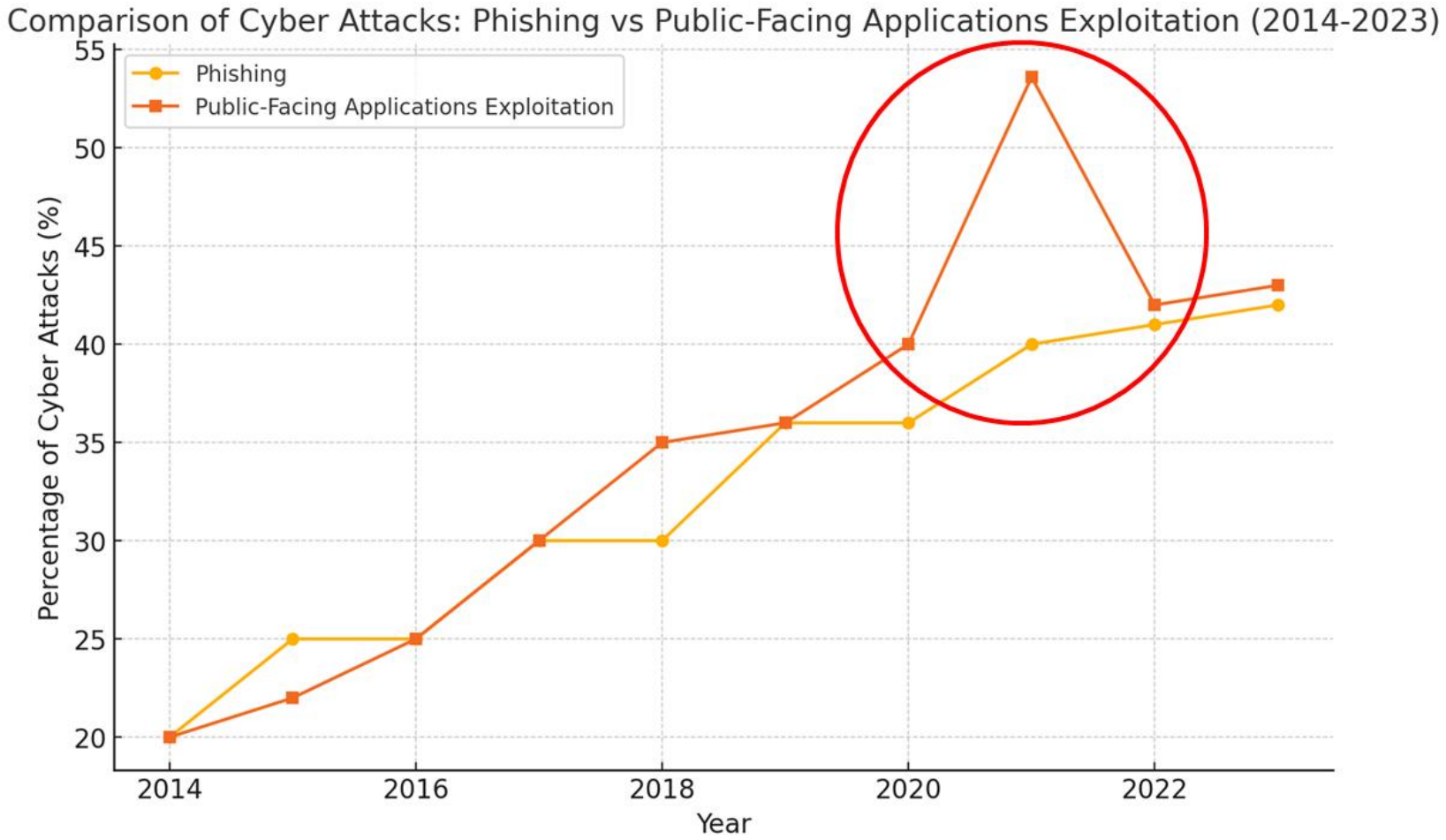
Cost of Cyber Crime w/ Emergence of AI

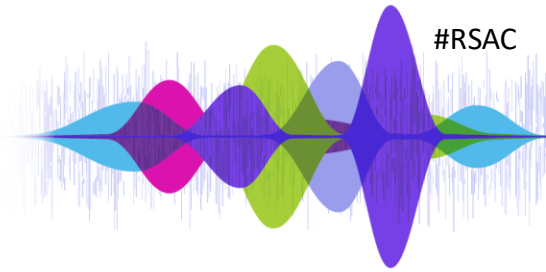


Successful Phishing w/ the Emergence of AI



Public-Facing Application Exploitation





Exploit development is super charged by AI tools.

Defenses are not.

RSAC 2025 Conference

Many Voices.
One Community.

The Real Threat of AI

Misinformation!

Fake News

FBI issues warning over two fake election videos

Fake Obama created using AI tool to make phoney speeches

Deepfakes porn has serious consequences

Experts fear face swapping tech could start an international showdown

Video forensic specialists are worried Deepfakes could have national security repercussions.

$18.5 Million Lost in a Crypto Scam: AI Voice Fraud's Growing Threat to Financial Security

Woman sentenced in case that sparked Springfield cat-eating rumours

LONG LIVE
THE REAL
KING

Deepfakes in the AI-Powered Threat Landscape

What is a deepfake?

A deepfake refers to a specific kind of synthetic media where a person in an image, video or audio file is swapped with another person's likeness.

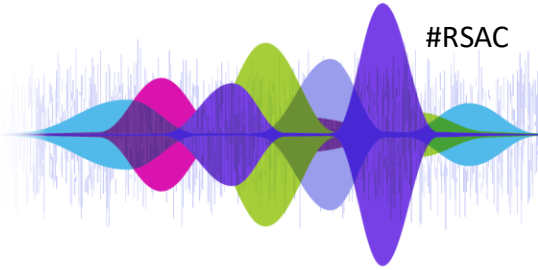
Use Cases / Examples

✅ Legitimate / Ethical Use Cases

- Entertainment & Film Production
- Accessibility & Language Translation
- Education & Training Simulations
- Personal & Creative Expression

❌ Malicious / Threat-Based Use Cases

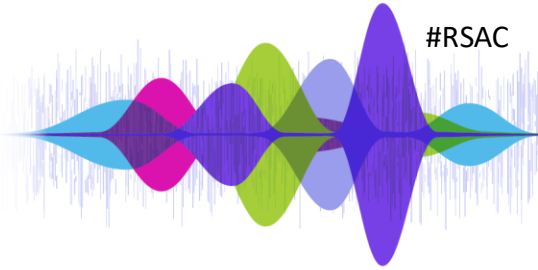
- Corporate Disinformation & Stock Manipulation
- Political & Election Manipulation
- Social Engineering & Cybercrime
- Voice Recognition Account Authorization
- Identity Theft or Celebrity Impersonation
- Misinformation in Conflict Zones



🔍 How to Identify Deepfakes

📼 Video Deepfakes

1. **Unnatural Facial Movements –** Inconsistent blinking, stiff expressions, or mismatched lip-syncing
2. **Skin Texture & Lighting -** Inconsistencies in lighting, shadows, or overly smooth skin
3. **Eye Reflections & Glasses -** Reflections may look off or absent; glasses may flicker or change shape unnaturally
4. **Asynchronous Audio & Video -** Voice may not match mouth movements or facial expressions
5. **Artifacts or Glitches -** Strange transitions, flickering, or blurring around the face or edges



🔍 How to Identify Deepfakes

🖼️ Image Deepfakes

1. **Asymmetry in Facial Features -** One eye might be slightly larger; ears may not match; teeth may look warped
2. **Weird Hands -** Hands with the wrong number of fingers or fingers turned the wrong direction.
3. **Blurred Background**

🔊 Audio Deepfakes

1. **Robotic or Monotone Voice -** May lack natural inflection, tone, or emotional variation
2. **Weird Pauses -** Strange speech rhythms, repeated intonations, or awkward pacing
3. **Background Noise Discrepancies -** The voice may sound too clean or not match the ambient sound of the video

RSAC 2025 Conference

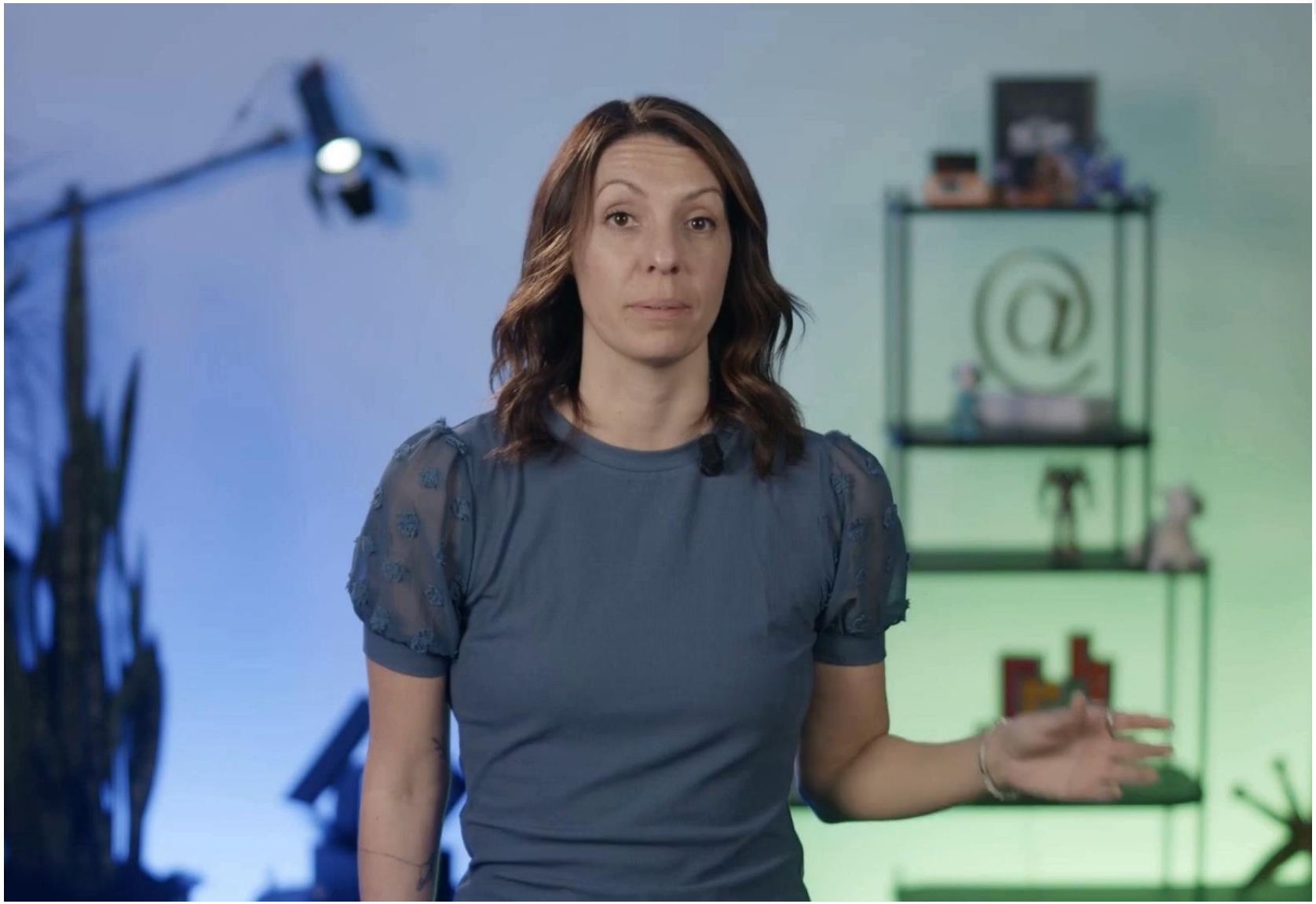
Let’s look at some examples

Which are real? Which are fake?

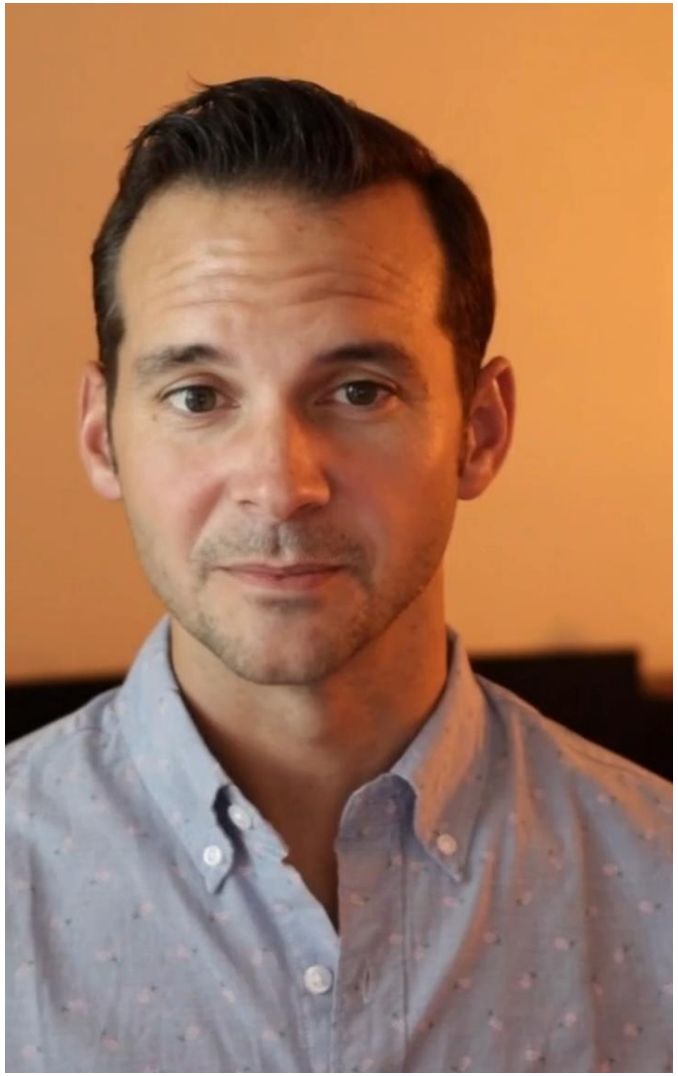
Many Voices.
One Community.

Can you spot it?

Real or Deepfake?



Real or Deepfake?



Real or Deepfake?

Real or Deepfake?

Real or Deepfake?

Real or Deepfake?

Real or Deepfake?

Real or Deepfake?

RSAC 2025 Conference

Lab Part One:
Time to make your own deepfake!

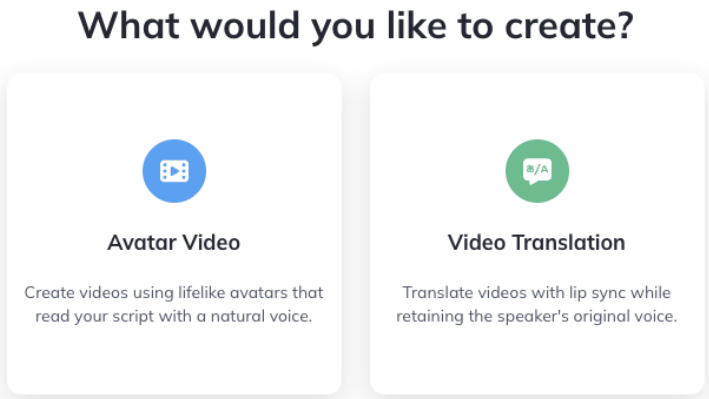
***Must have a webcam or phone camera or partner with someone who does**

Many Voices.
One Community.

https://www.heygen.com/

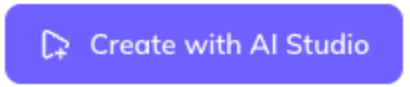
- Create an account
  - This is a free account, and you will not be charged
- Create Video

- Create Avatar Video


What would you like to create?
Avatar Video
Create videos using lifelike avatars that read your script with a natural voice.
Video Translation
Translate videos with lip sync while retaining the speaker's original voice.

https://www.heygen.com/

- Record with Webcam

- Take 90 seconds to follow the prompts to create your Avatar
  - ***If you have airpods/headphones here, use them for audio/echo
- Avatar will take ~5 minutes to render
- Choose newly created avatar and then “Create with AI Studio” 

- Click Script on the Left and paste in the following:
  - Hi there! My name is <your name.> I’m attending the RSAC 2025 conference in San Francisco. This workshop learning about deepfakes is amazing!

https://www.heygen.com/

- Click the play button to see how the playback reacts
- Click the ellipsis(…) to modify voice settings (Speed, Model, Style Exaggeration) and then click play as you see fit
- When you are happy with the playback, click submit!
- Fill in Filename “RSAC-D” (D for Deepfake), and click ”save”
- Videos will take a few minutes to render. It will says “ready” when complete.
- Click play to watch your newly created deepfake.
- Finally, download the file for “distribution”

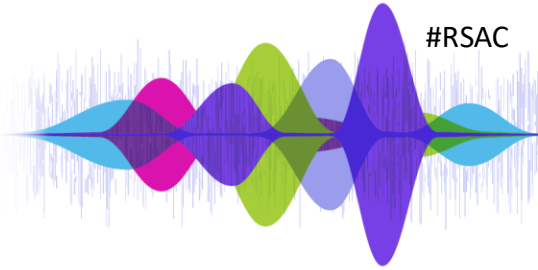
RSAC 2025 Conference

Lab Part Two: Record a real video

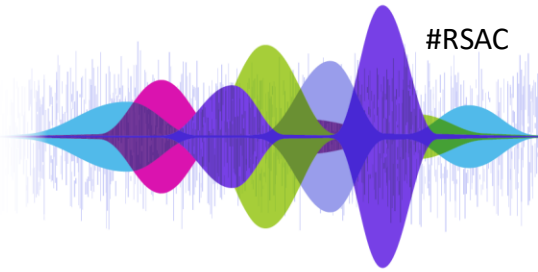
Walkthrough on both Mac, PC and iPhone

Many Voices.
One Community.

Record on your phone *Easiest*



- iPhone or Android
  - Open your camera app, record a video and airdrop/send to your computer
- Use the same script:
  - Hi there! My name is <your name.> I’m attending the RSA 2025 conference in San Francisco. This workshop learning about deepfakes is amazing!



Record on your laptop

- iMovie (mac)
- Quicktime player (PC)
- Zoom
- Streamlabs

- Use the same script:
  - Hi there! My name is <your name.> I'm attending the RSAC 2025 conference in San Francisco. This workshop learning about deepfakes is amazing!

RSAC 2025 Conference

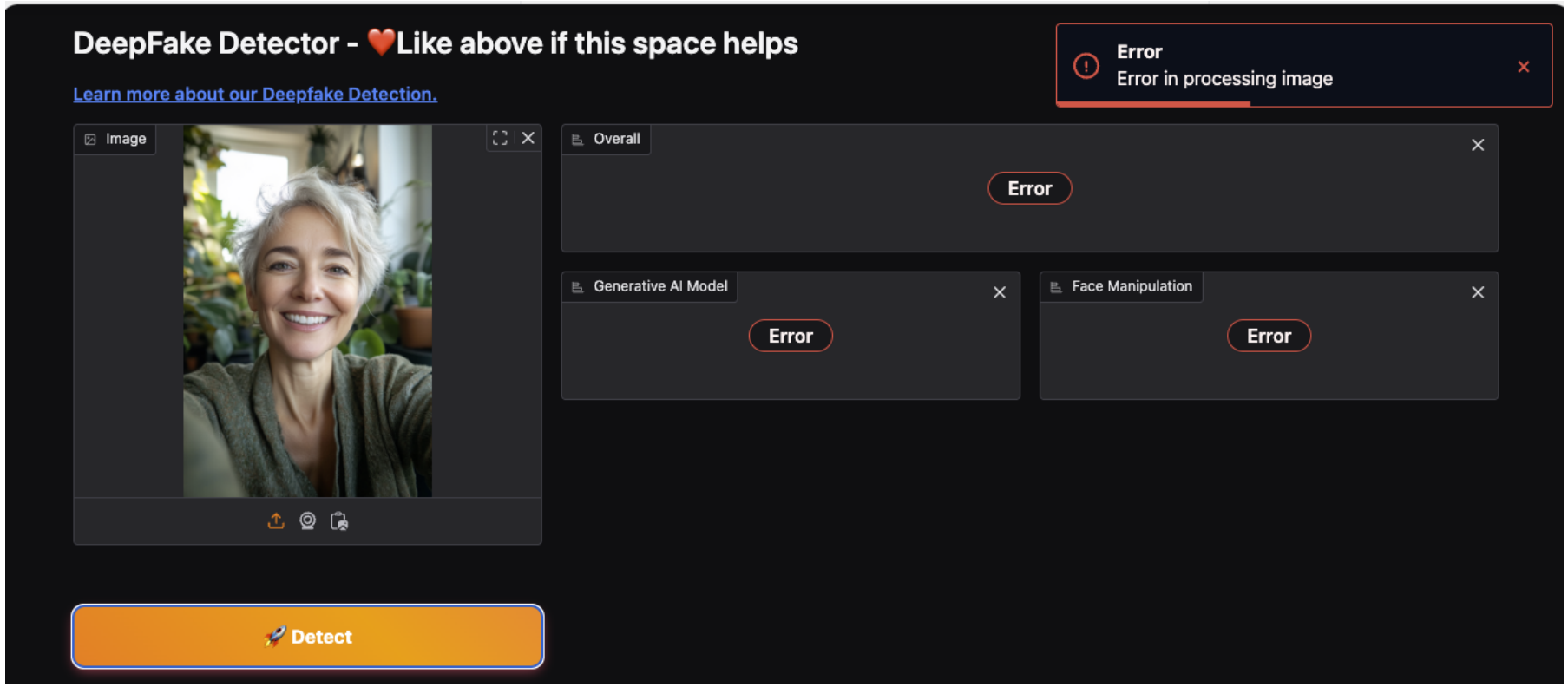
Lab Part Three: Deepfake Detection

Walkthrough on both Mac, PC and iPhone

Many Voices.
One Community.

FaceOnLive – Deepfake Detector

Image deepfake

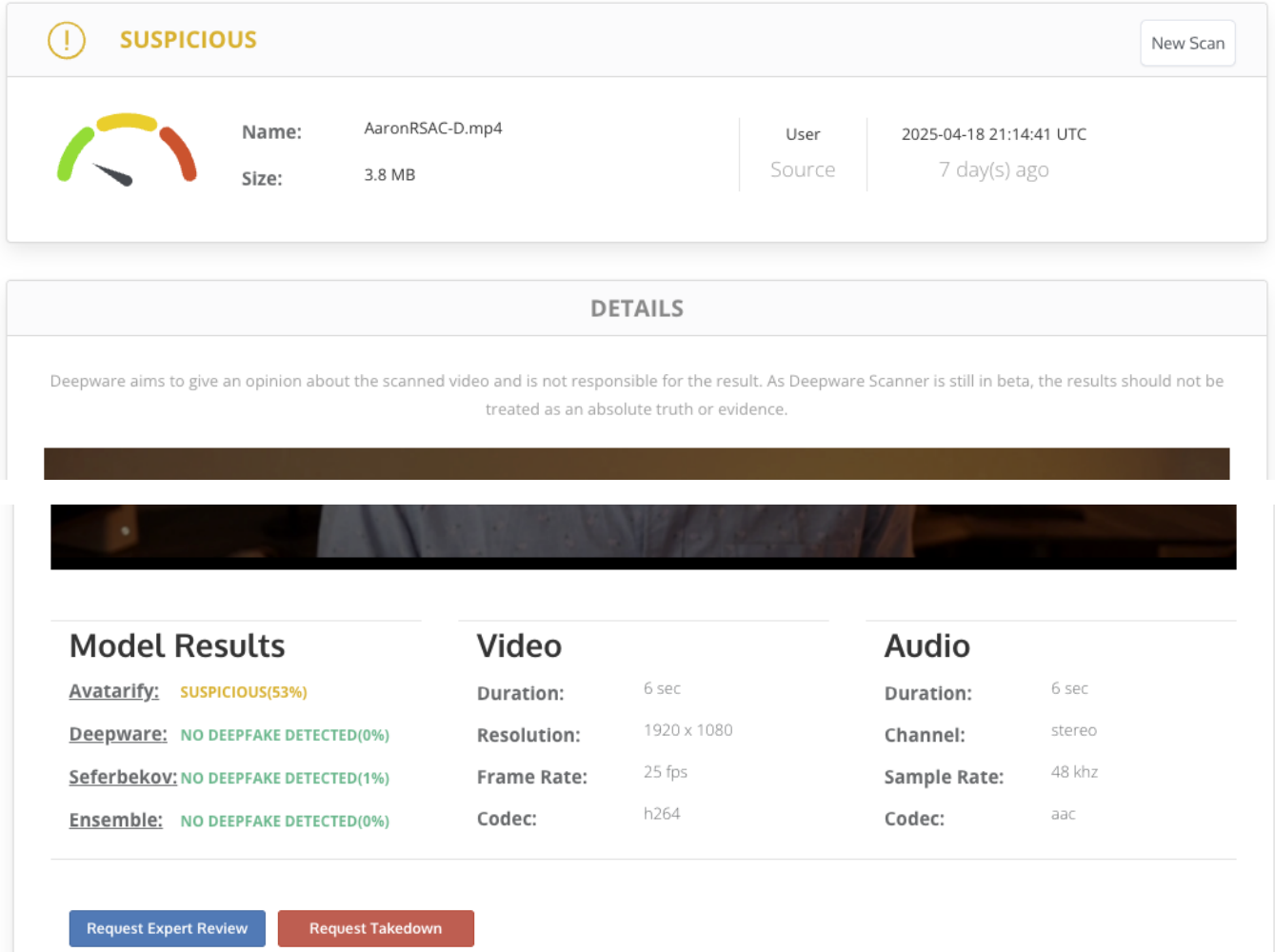


Deepware Scanner

Aaron RSA Deepfake

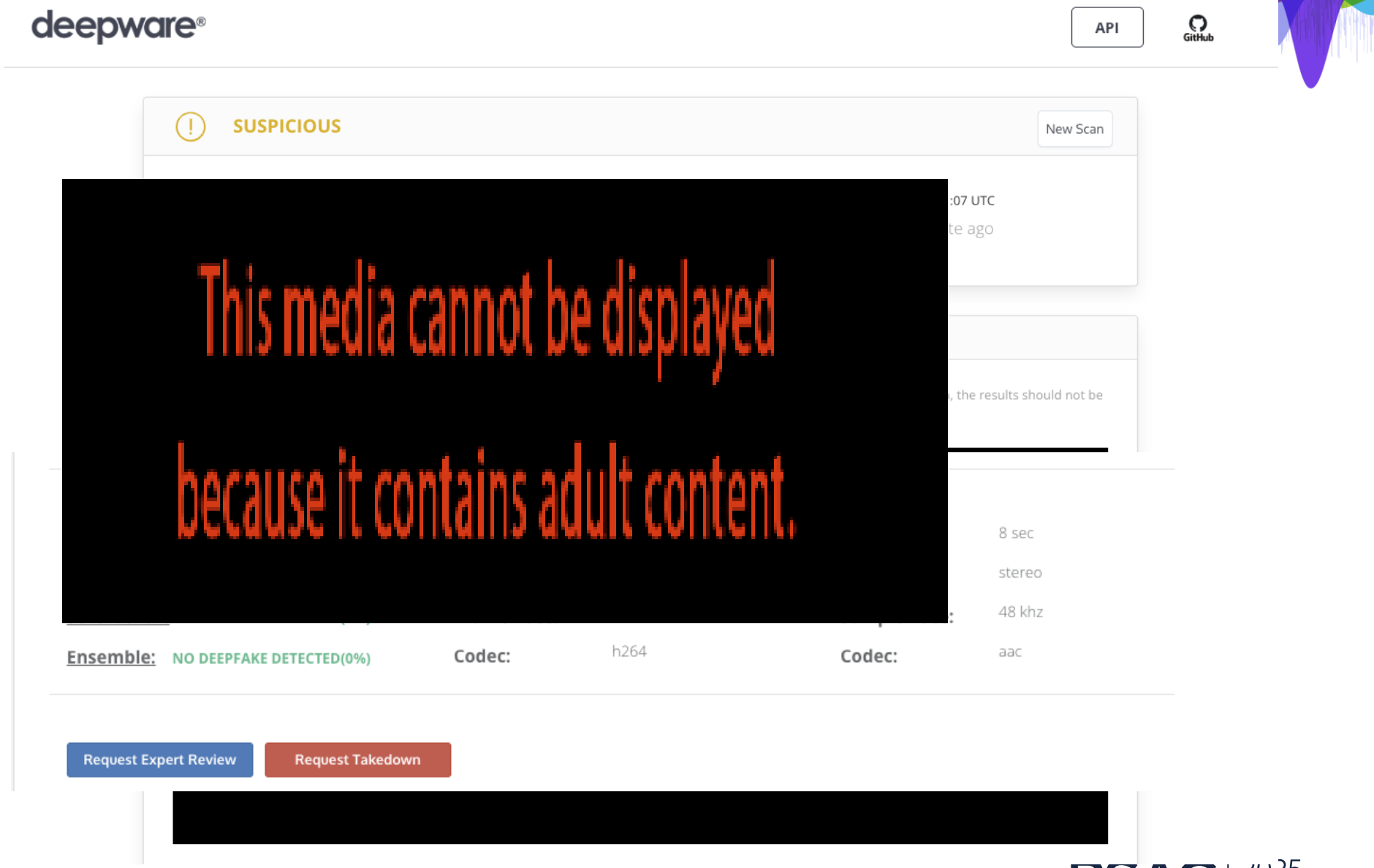
deepware®

API GitHub



Deepware Scanner

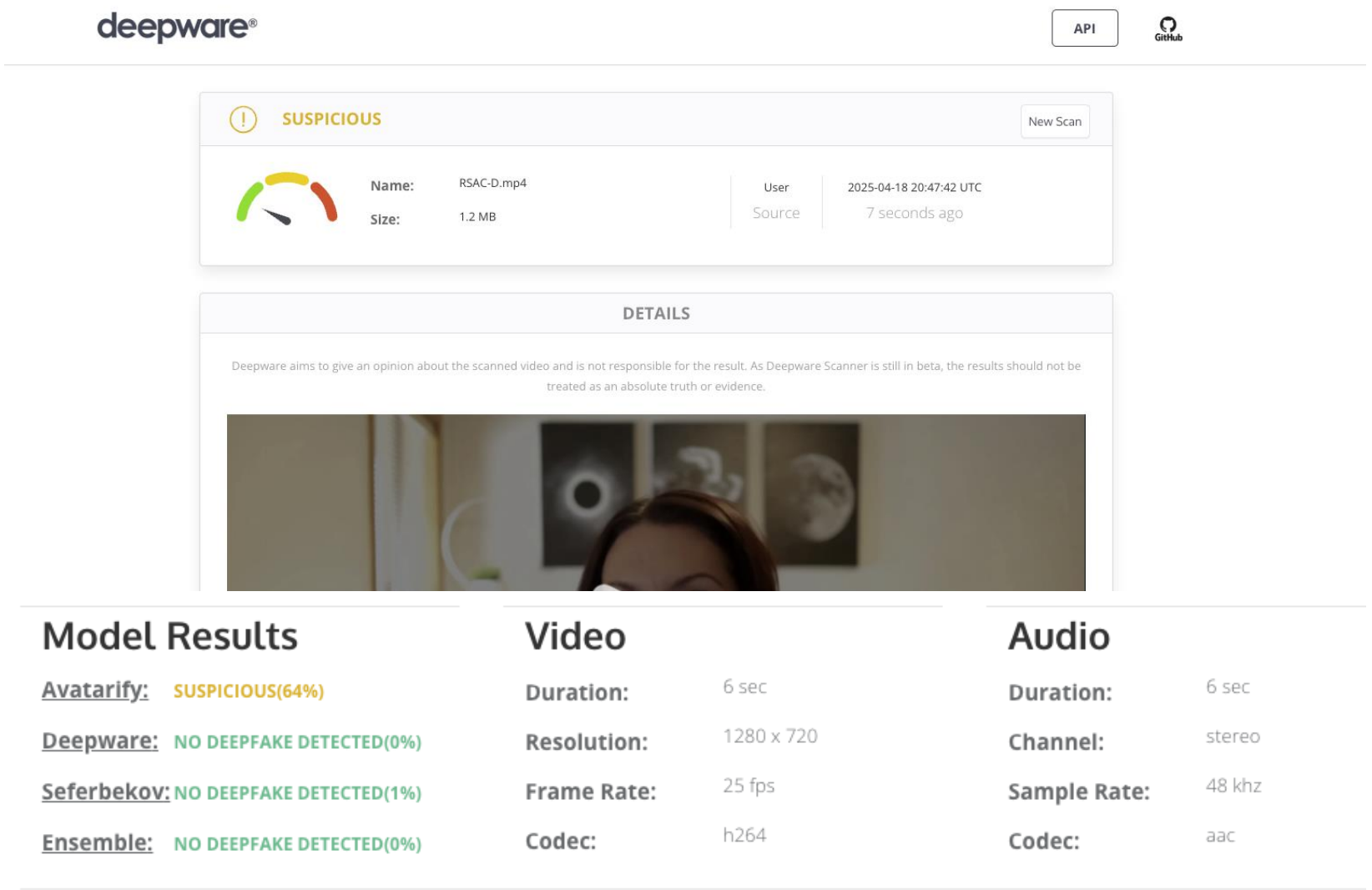
Aaron Github Deepfake



#RSAC
RSAC 2025 Conference

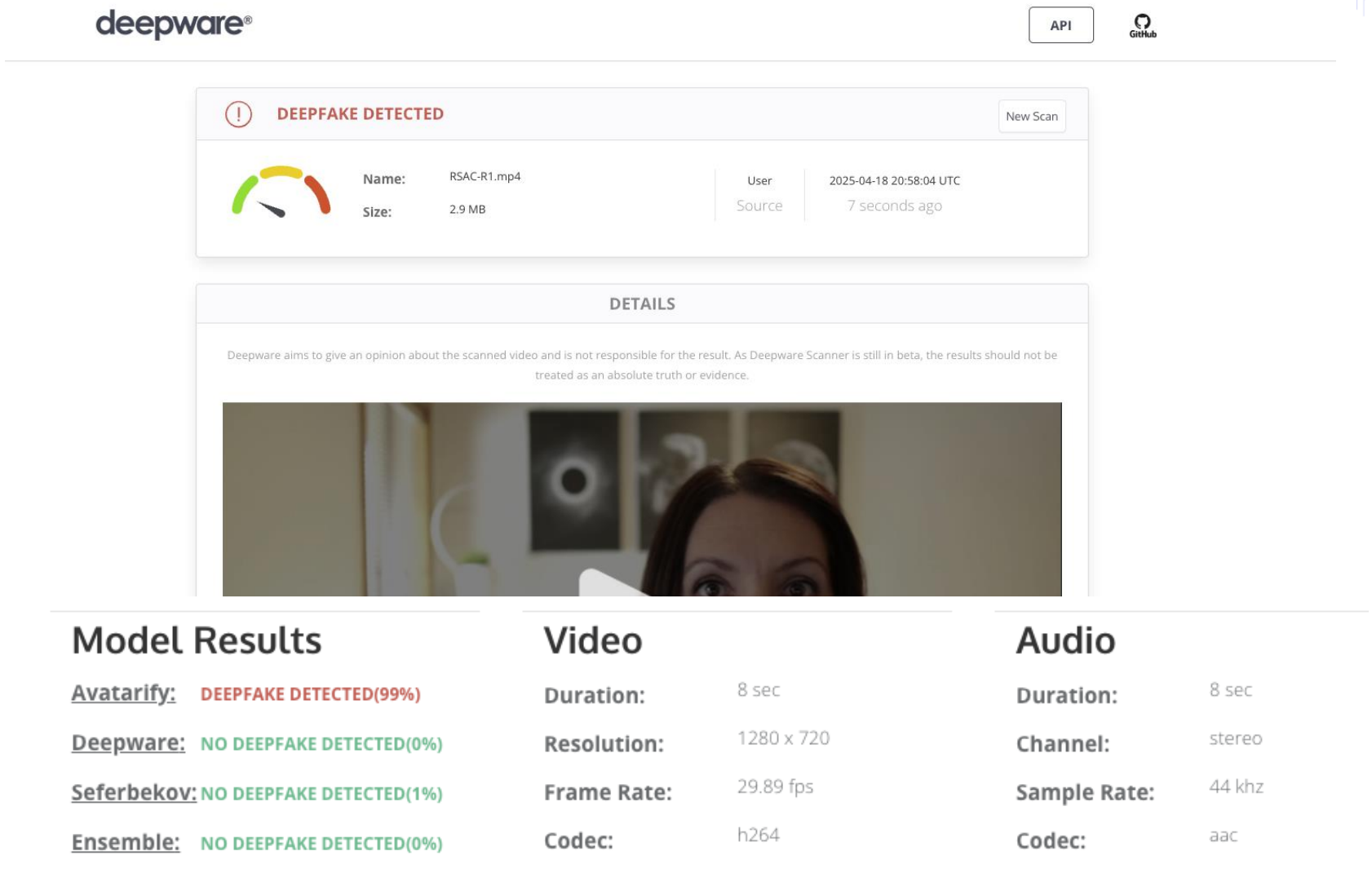
Deepware Scanner

Bri RSA
Deepfake



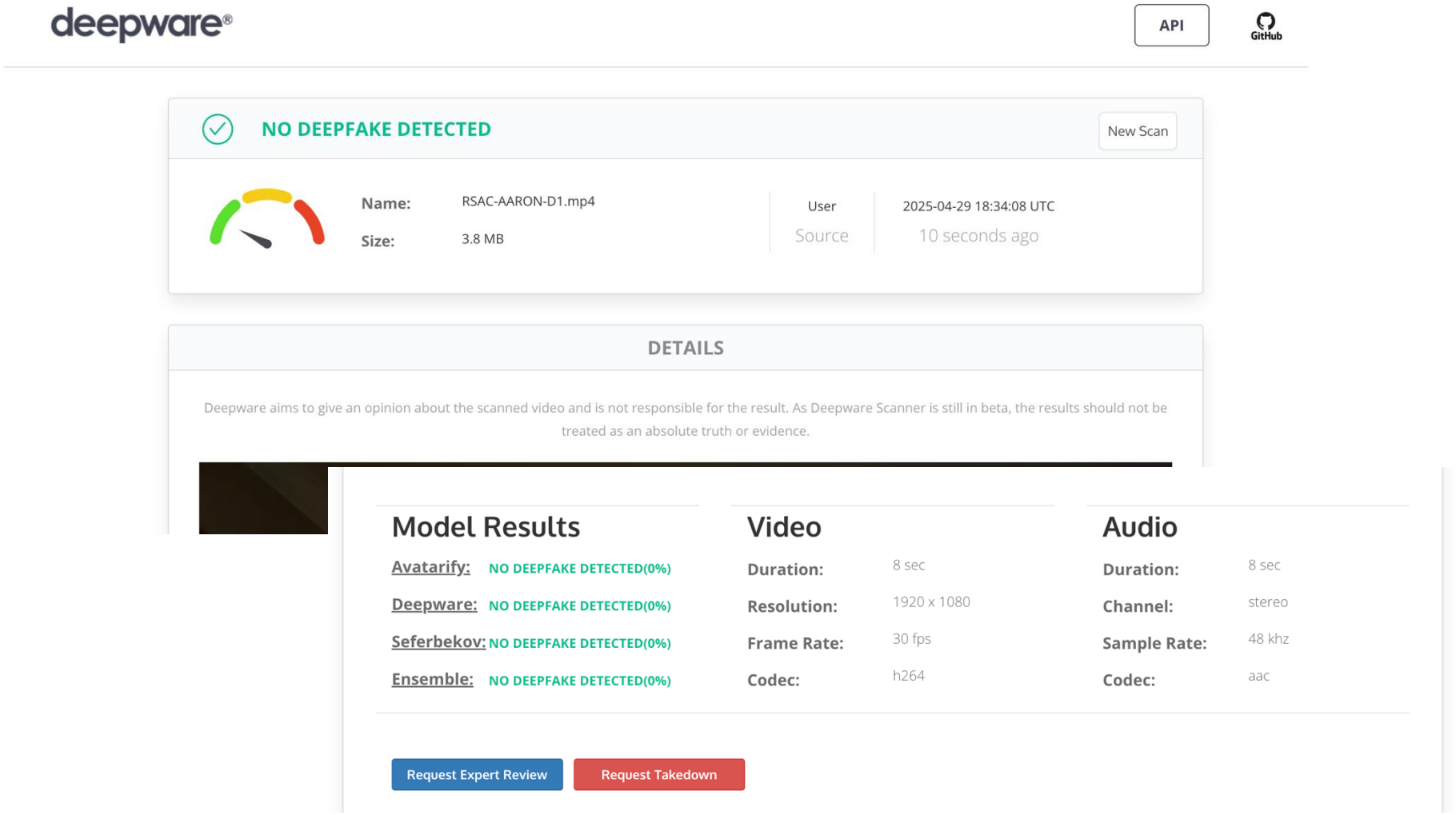
Deepware Scanner

Bri RSA
Real



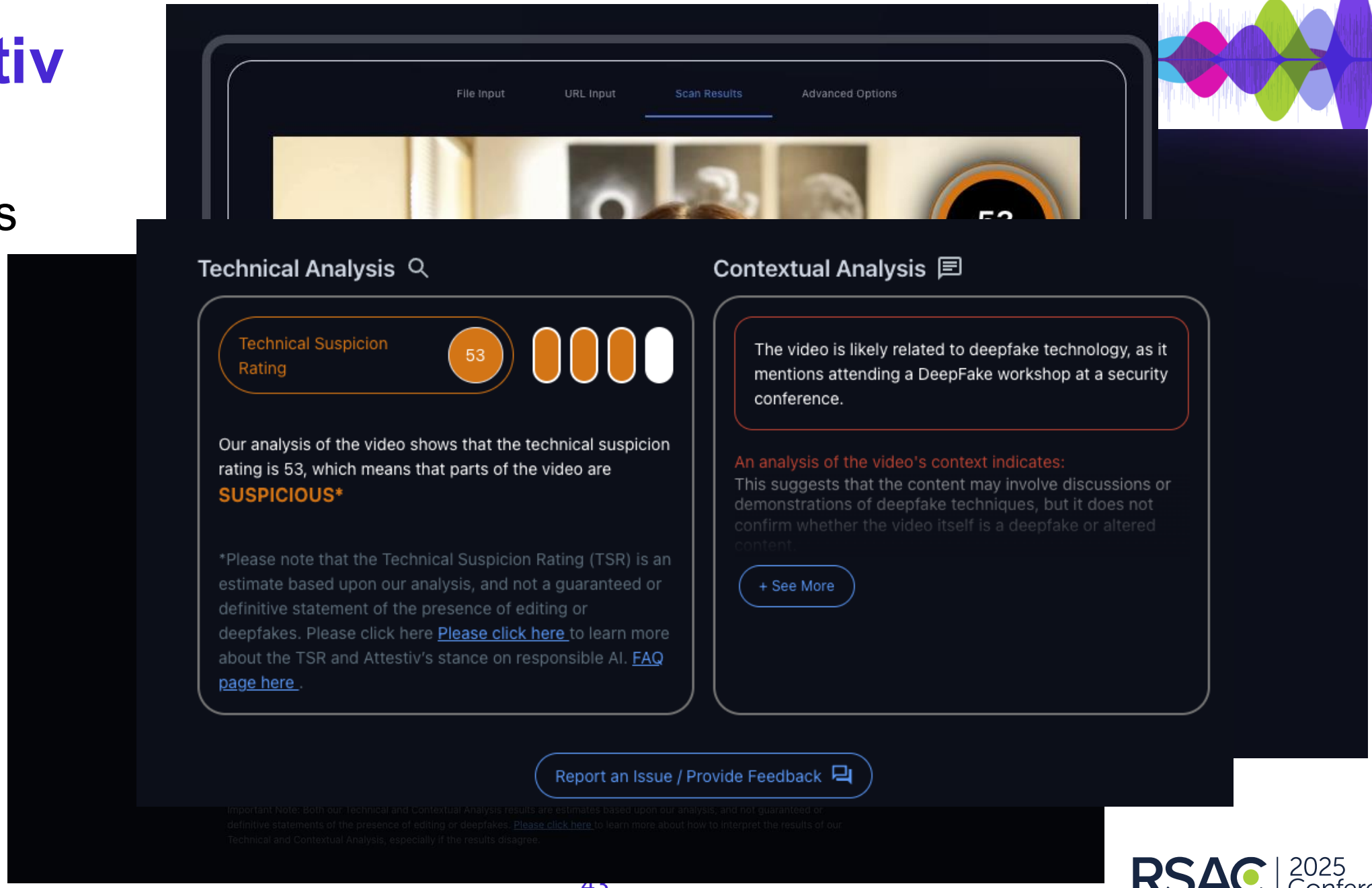
Deepware Scanner

Aaron Deepfake



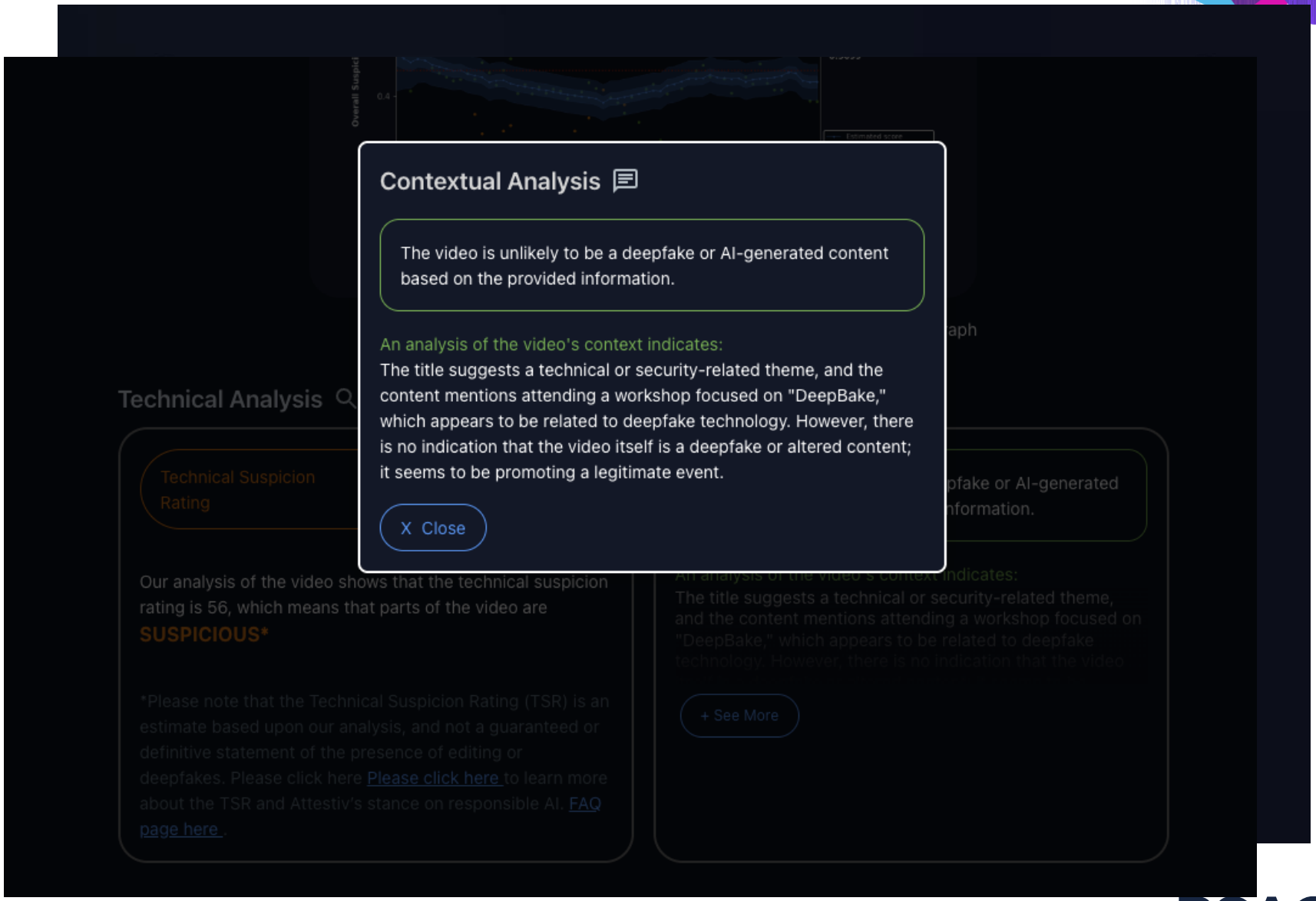
Attestiv

RSA Deepfakes


File Input
URL Input
Scan Results
Advanced Options
Technical Analysis
Technical Suspicion Rating
53
Our analysis of the video shows that the technical suspicion rating is 53, which means that parts of the video are SUSPICIOUS*
*Please note that the Technical Suspicion Rating (TSR) is an estimate based upon our analysis, and not a guaranteed or definitive statement of the presence of editing or deepfakes. Please click here Please click here to learn more about the TSR and Attestiv's stance on responsible AI. FAQ page here .
Contextual Analysis
The video is likely related to deepfake technology, as it mentions attending a DeepFake workshop at a security conference.
An analysis of the video's context indicates:
This suggests that the content may involve discussions or demonstrations of deepfake techniques, but it does not confirm whether the video itself is a deepfake or altered content.
+ See More
Report an Issue / Provide Feedback
Important Note: Both our Technical and Contextual Analysis results are estimates based upon our analysis, and not guaranteed or definitive statements of the presence of editing or deepfakes. Please click here to learn more about how to interpret the results of our Technical and Contextual Analysis, especially if the results disagree.

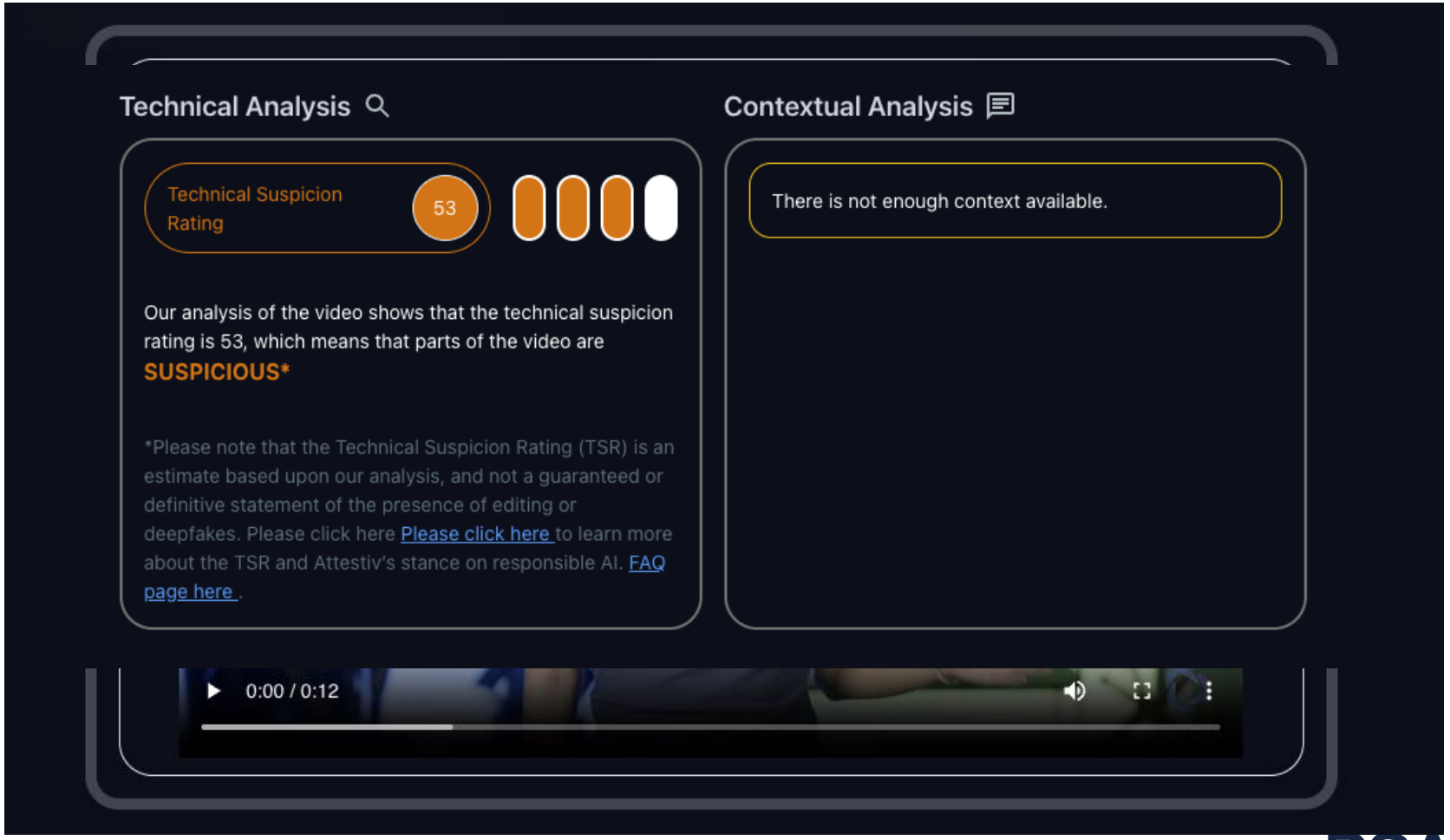
Attestiv

Bri RSA
Real


Contextual Analysis
The video is unlikely to be a deepfake or AI-generated content based on the provided information.
An analysis of the video's context indicates:
The title suggests a technical or security-related theme, and the content mentions attending a workshop focused on "DeepBake," which appears to be related to deepfake technology. However, there is no indication that the video itself is a deepfake or altered content; it seems to be promoting a legitimate event.
X Close
Technical Analysis
Technical Suspicion Rating
Our analysis of the video shows that the technical suspicion rating is 56, which means that parts of the video are SUSPICIOUS*
*Please note that the Technical Suspicion Rating (TSR) is an estimate based upon our analysis, and not a guaranteed or definitive statement of the presence of editing or deepfakes. Please click here Please click here to learn more about the TSR and Attestiv's stance on responsible AI. FAQ page here .
+ See More

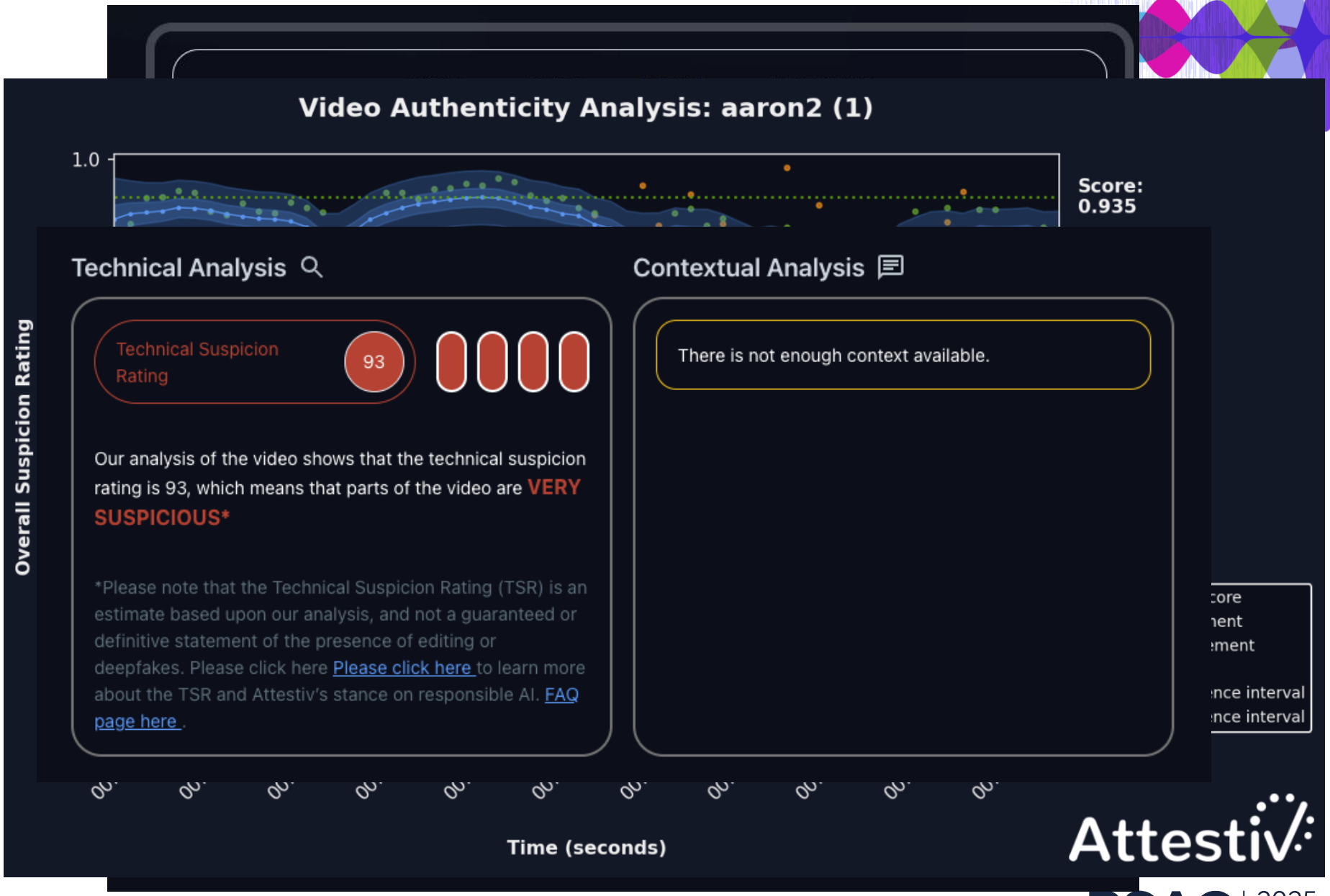
Attestiv

Bri Security Threat Deepfake


Technical Analysis
Contextual Analysis
Technical Suspicion Rating
53
There is not enough context available.
Our analysis of the video shows that the technical suspicion rating is 53, which means that parts of the video are SUSPICIOUS*
*Please note that the Technical Suspicion Rating (TSR) is an estimate based upon our analysis, and not a guaranteed or definitive statement of the presence of editing or deepfakes. Please click here Please click here to learn more about the TSR and Attestiv's stance on responsible AI. FAQ page here .
0:00 / 0:12

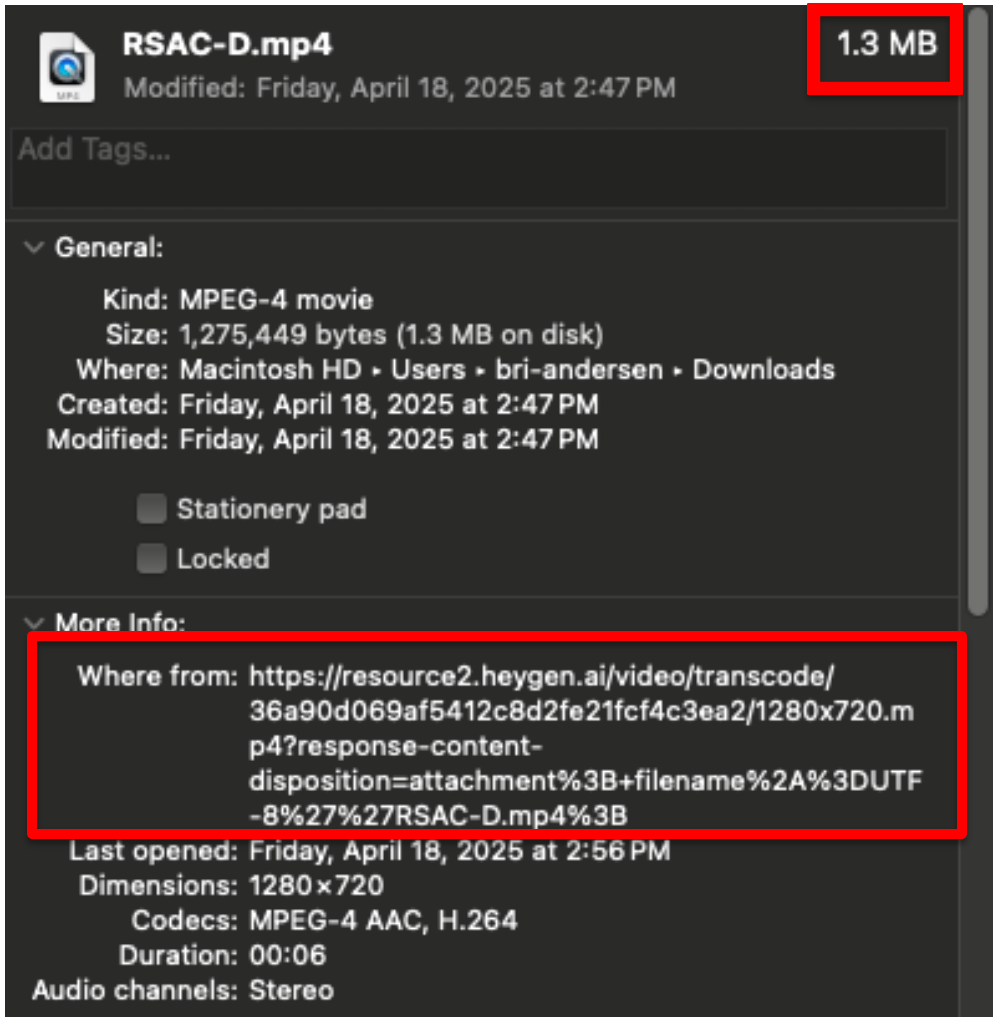
Attestiv

Aaron Github Deepfake

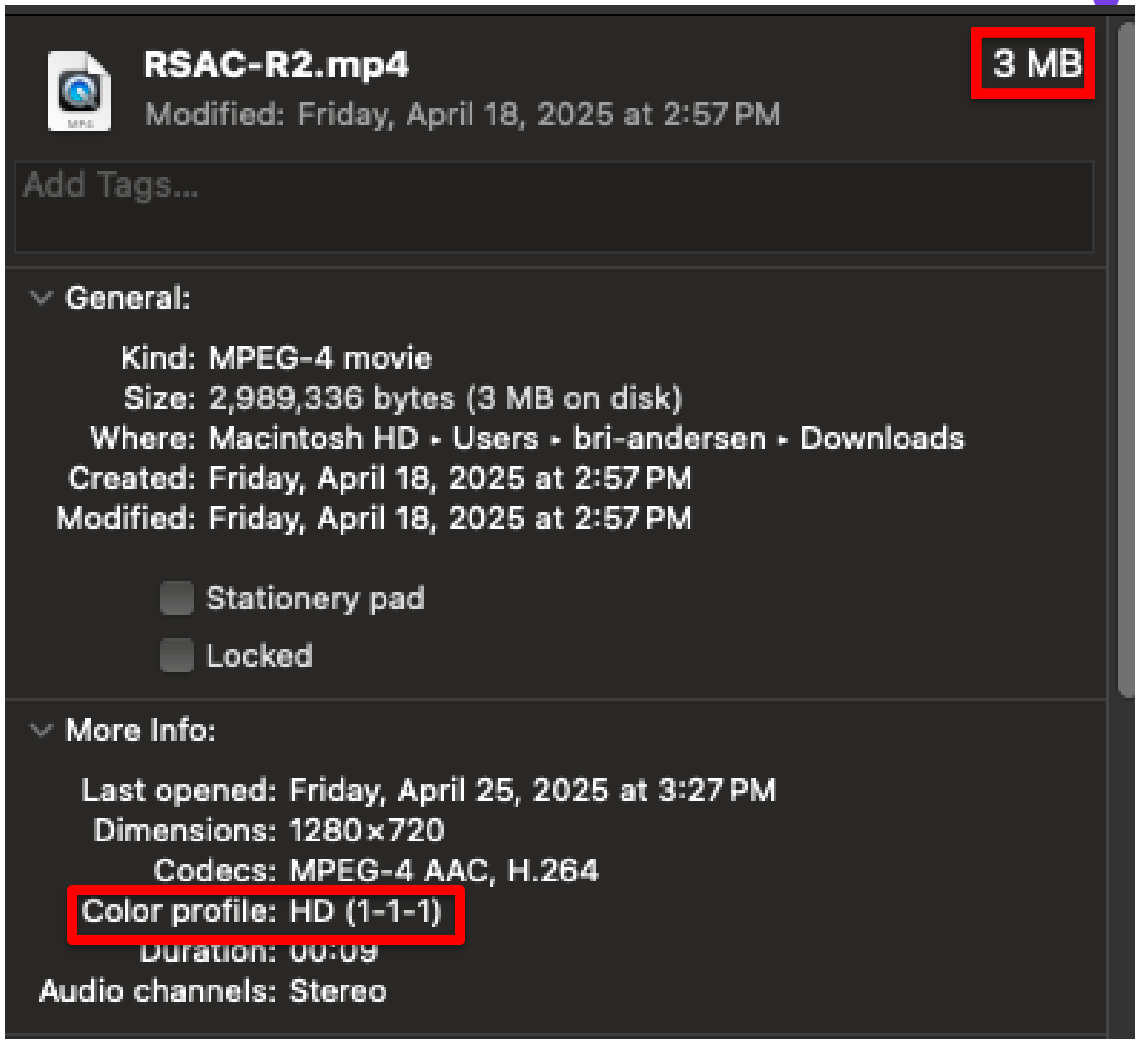

Video Authenticity Analysis: aaron2 (1)
1.0
Score:
0.935
Technical Analysis
Contextual Analysis
Technical Suspicion Rating
93
There is not enough context available.
Our analysis of the video shows that the technical suspicion rating is 93, which means that parts of the video are VERY SUSPICIOUS*
*Please note that the Technical Suspicion Rating (TSR) is an estimate based upon our analysis, and not a guaranteed or definitive statement of the presence of editing or deepfakes. Please click here Please click here to learn more about the TSR and Attestiv's stance on responsible AI. FAQ page here .
Overall Suspicion Rating
Time (seconds)
Attestiv

Metadata Comparison

RSAC-D

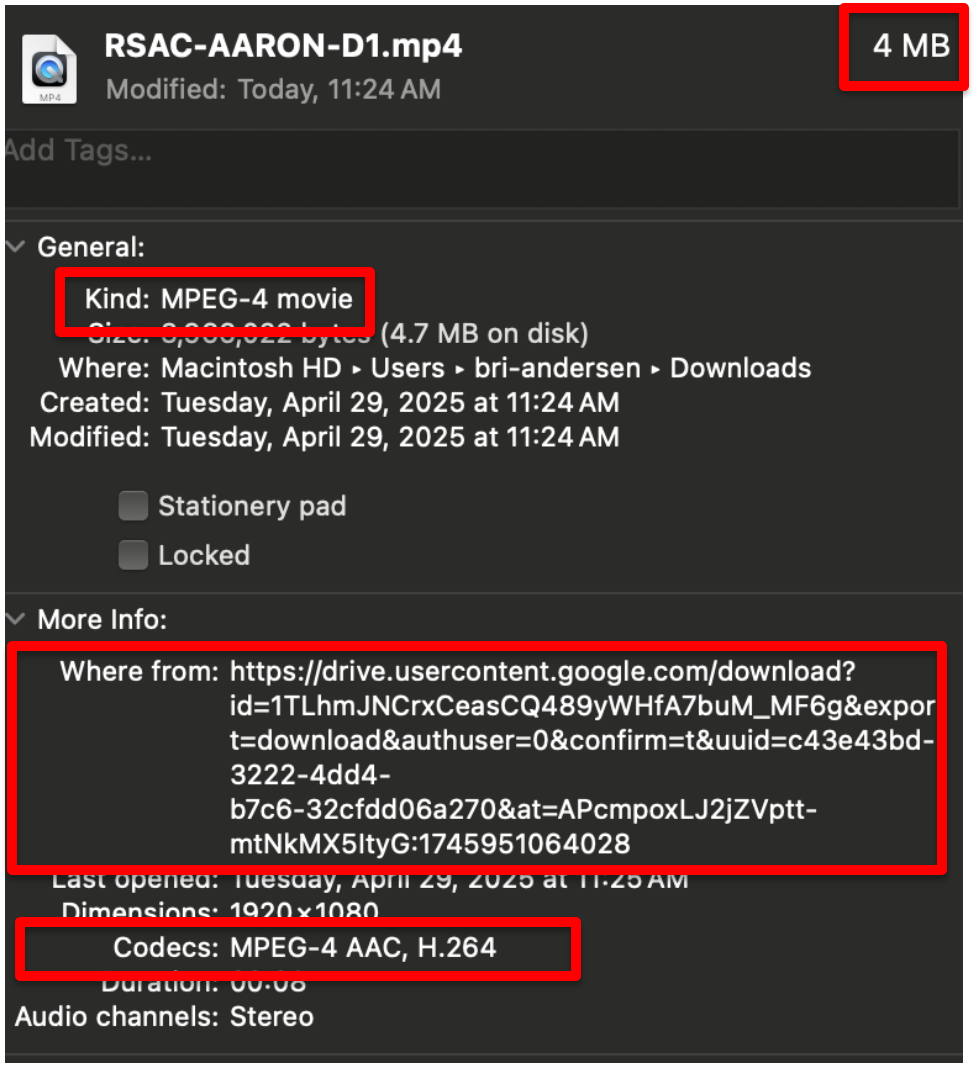

RSAC-D.mp4
1.3 MB
Modified: Friday, April 18, 2025 at 2:47 PM
Add Tags...
General:
Kind: MPEG-4 movie
Size: 1,275,449 bytes (1.3 MB on disk)
Where: Macintosh HD ▸ Users ▸ bri-andersen ▸ Downloads
Created: Friday, April 18, 2025 at 2:47 PM
Modified: Friday, April 18, 2025 at 2:47 PM
Stationery pad
Locked
More Info:
Where from: https://resource2.heygen.ai/video/transcode/36a90d069af5412c8d2fe21fcf4c3ea2/1280x720.mp4?response-content-disposition=attachment%3B+filename%2A%3DUTF-8%27%27RSAC-D.mp4%3B
Last opened: Friday, April 18, 2025 at 2:56 PM
Dimensions: 1280×720
Codecs: MPEG-4 AAC, H.264
Duration: 00:06
Audio channels: Stereo

RSAC-R

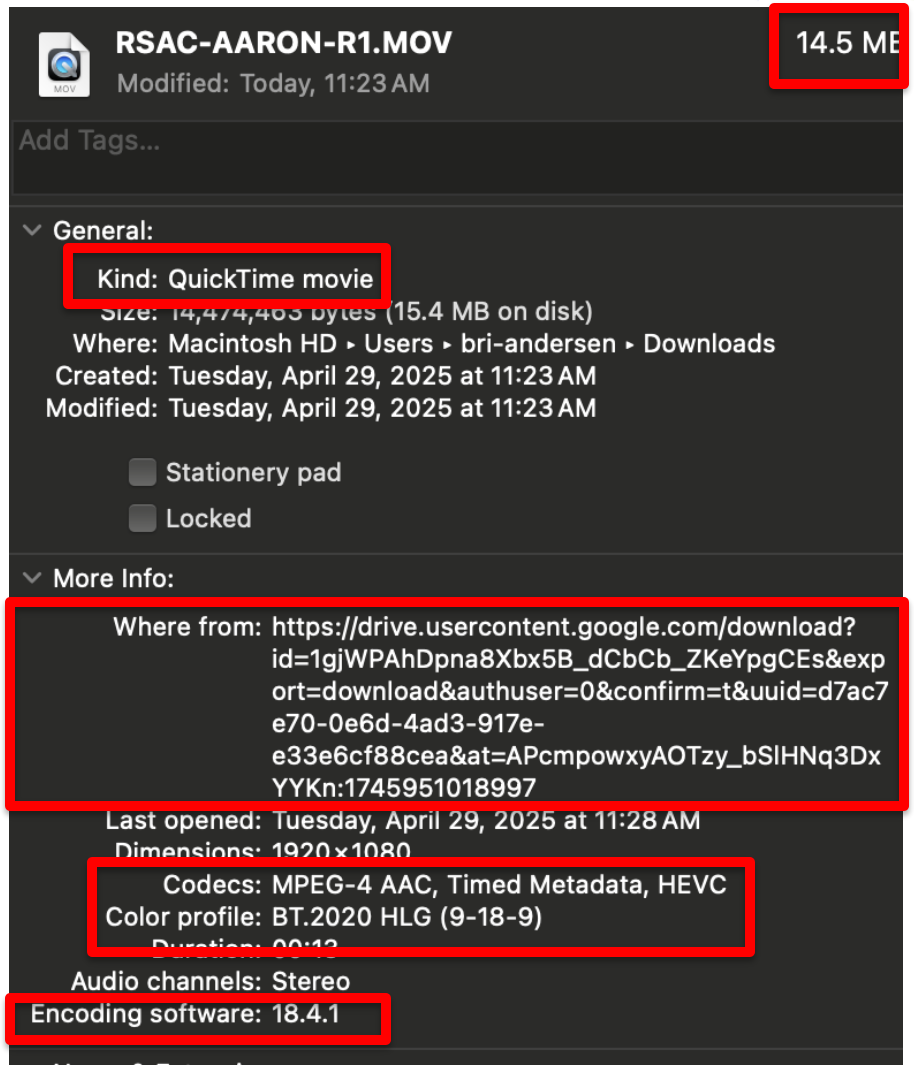

RSAC-R2.mp4
3 MB
Modified: Friday, April 18, 2025 at 2:57 PM
Add Tags...
General:
Kind: MPEG-4 movie
Size: 2,989,336 bytes (3 MB on disk)
Where: Macintosh HD ▸ Users ▸ bri-andersen ▸ Downloads
Created: Friday, April 18, 2025 at 2:57 PM
Modified: Friday, April 18, 2025 at 2:57 PM
Stationery pad
Locked
More Info:
Last opened: Friday, April 25, 2025 at 3:27 PM
Dimensions: 1280×720
Codecs: MPEG-4 AAC, H.264
Color profile: HD (1-1-1)
Audio channels: Stereo

Metadata Comparison

RSAC-AARON-D1



RSAC-AARON-R1



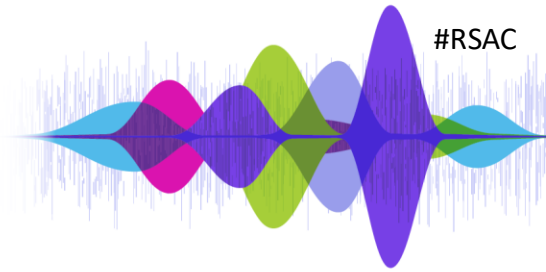
RSAC 2025 Conference

Conclusion + What's next?

Mitigating Deepfake Threats in the Future

Many Voices.
One Community.

How to Mitigate Deepfakes



Digital Provenance & Watermarking

- **Use tools like Content Credentials (C2PA)** to verify authenticity at the point of creation.
- Embed metadata or blockchain-based timestamps that confirm source and integrity.

User & Employee Awareness Training

- Educate teams to spot suspicious content, especially in social engineering contexts (e.g., deepfake CEO on Zoom).
- Include examples in cybersecurity awareness programs.

Multi-factor Validation!

- confirmation.

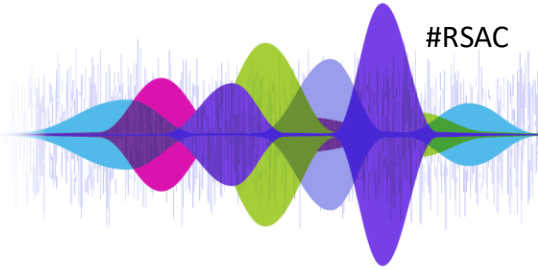
5. Access Controls for Sensitive Media

- Limit access to high-quality video/audio of executives and VIPs to prevent training data leaks.
- Scrub or watermark public appearances with anti-AI training techniques (e.g., adversarial filters).

RSAC 2025 Conference

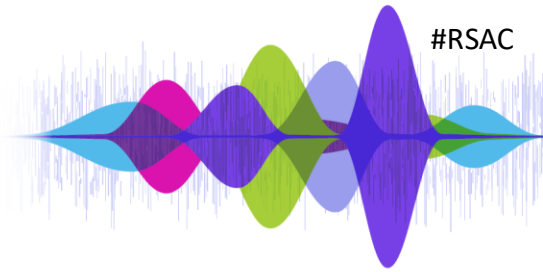
🎯 "Technology may create the deepfake, but **awareness**, **verification**, and **detection** are how we fight back."

Apply What You Have Learned Today



- Next week you should:
  - Spread awareness! Share a video using your AI avatar. Tell your coworkers, friends and families the ways to identify a deepfake video.
- In the first three months following this presentation you should:
  - Validation! Talk about Multi-factor Validation with your network. Encourage your security and IT teams to utilize deepfake identification tools.
- Within six months you should:
  - Detection! Continue to use AI-creation as well as Deepfake detection tools. These technologies are brand new and escalating in their capability quickly! If we think like the adversary and use the same tools as the adversary, it takes away their power!

Lessons Learned – What's Next?



- What threat do deepfakes pose to me and my organization?
  - How does that affect risk?
  - What is your current policy for risk within that same category?
- Do deepfakes require a change in the way that risk is handled?
- Should AI be used to fight AI?

RSAC 2025 Conference

Many Voices.
One Community.

Thank you!

Questions?